Spam Detection Using Naive Bayes Classifier

Created by: Jangita Takeshwar

Introduction

This Python script builds a spam detection system using the Naive Bayes classifier. It processes a dataset of labeled emails, applies text preprocessing techniques, and uses TF-IDF vectorization to convert text into numerical features. The trained model then predicts whether a given email is spam or not.

Importing Libraries

The following libraries are used in the script:

- pandas and numpy: Used for data manipulation.

- re and string: Used for text preprocessing.

- matplotlib.pyplot: Used for data visualization.

- sklearn.model_selection.train_test_split: Splits the dataset into training and testing sets.

- sklearn.feature_extraction.text.TfidfVectorizer: Converts text into numerical TF-IDF features.

- sklearn.naive_bayes.MultinomialNB: The Naive Bayes classifier used for text classification.

- sklearn.metrics.accuracy_score and classification_report: Evaluate model performance.

Loading and Inspecting the Dataset

The dataset is read from a CSV file, and the first few rows are displayed using the head() function.

Data Cleaning and Preprocessing

The dataset is preprocessed by selecting the necessary columns, renaming them, converting labels to binary values, and applying text cleaning techniques such as removing punctuation and numbers.

Splitting the Dataset

The dataset is split into training (80%) and testing (20%) sets using train_test_split, ensuring reproducibility with a fixed random state.

Feature Extraction with TF-IDF

TF-IDF vectorization is used to convert text into a numerical format. Stop words are removed to improve efficiency.

Training the Naive Bayes Model

A Multinomial Naive Bayes classifier is initialized and trained using the TF-IDF features.

Making Predictions and Evaluating the Model

The trained model is tested on the test dataset. Accuracy and a classification report are generated to assess performance.

Visualizing Data Distribution

A bar chart is plotted to show the distribution of spam and ham emails in the dataset.

Function to Predict New Emails

A function is defined to predict whether a given email is spam or not based on the trained model. An example email is provided for testing.

Conclusion

This script successfully implements a spam detection system using the Naive Bayes classifier. It preprocesses the dataset, extracts TF-IDF features, trains a model, evaluates its performance, and includes a function for real-time spam classification.